Lecture Notes in Artificial Intelligence 4343

Edited by J. G. Carbonell and J. Siekmann

Subseries of Lecture Notes in Computer Science

Christian Müller (Ed.)

Speaker Classification I

Fundamentals, Features, and Methods

Springer

Series Editors

Jaime G. Carbonell, Carnegie Mellon University, Pittsburgh, PA, USA
Jörg Siekmann, University of Saarland, Saarbrücken, Germany

Volume Editor

Christian Müller
International Computer Science Institute
1947 Center Street, Berkeley, CA 94704, USA
E-mail: cmueller@icsi.berkeley.edu

Library of Congress Control Number: 2007932293

CR Subject Classification (1998): I.2.7, I.2.6, H.5.2, H.5, I.4-5

LNCS Sublibrary: SL 7 – Artificial Intelligence

ISSN	0302-9743
ISBN-10	3-540-74186-0 Springer Berlin Heidelberg New York
ISBN-13	978-3-540-74186-2 Springer Berlin Heidelberg New York

Springer is a part of Springer Science+Business Media

springer.com

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 12107810 06/3180 5 4 3 2 1 0

Preface

> "As well as conveying a message in words and sounds, the speech signal carries information about the speaker's own anatomy, physiology, linguistic experience and mental state. These speaker characteristics are found in speech at all levels of description: from the spectral information in the sounds to the choice of words and utterances themselves."

The best way to introduce this textbook is by using the words Volker Dellwo and his colleagues had chosen to begin their chapter "How Is Individuality Expressed in Voice?" While they use this statement to motivate the introductory chapter on speech production and the phonetic description of speech, it constitutes a framework of the entire book as well: What characteristics of the speaker become manifest in his or her voice and speaking behavior? Which of them can be inferred from analyzing the acoustic realizations? What can this information be used for? Which methods are the most suitable for diversified problems in this area of research? How should the quality of the results be evaluated?

Within the scope of this book the term *speaker classification* is defined as assigning a given speech sample to a particular class of speakers. These classes could be Women vs. Men, Children vs. Adults, Natives vs. Foreigners, etc. *Speaker recognition* is considered as being a sub-field of speaker classification in which the respective class has only one member (Speaker vs. Non-Speaker). Since in the engineering community this sub-field is explored in more depth than others covered by the book, many of the articles focus on speaker recognition. Nevertheless, the findings are discussed in the context of the broader notion of speaker classification where feasible.

The book is organized in two volumes. Volume I encompasses more general and overview-like articles which contribute to answering a subset of the questions above: Besides Dellwo and coworkers' introductory chapter, the "Fundamentals" part also includes a survey by David Hill, who addresses past and present speaker classification issues and outlines a potential future progression of the field.

The subsequent part is concerned with the multitude of candidate speaker "Characteristics." Tanja Schulz describes "why it is desirable to automatically derive particular speaker characteristics from speech" and focuses on language, accent, dialect, ideolect, and sociolect. Ulrike Gut investigates "how speakers can be classified into native and non-native speakers of a language on the basis of acoustic and perceptually relevant features in their speech" and compiles a list of the most salient acoustic properties of foreign accent. Susanne Schötz provides a survey about speaker age, covering the effects of ageing on the speech production mechanism, the human ability of perceiving speaker age, as well as its automatic recognition. John Hansen and Sanjay Patil "consider a range of issues associated with analysis, modeling, and recognition of speech under stress." Anton Batliner and Richard Huber address the problem of emotion classification focusing on the specific phenomenon of irregular phonation or laryngealization and thereby point out the inherent problem of speaker-dependency, which relates the problems of speaker identification and emotion recognition with each other. The juristic implications of acquiring knowledge about the speaker on the basis of his or her speech in the context of emotion recognition is addressed by Erik Eriksson and his co-authors, discussing, "inter alia, assessment of emotion in others, witness credibility, forensic investigation, and training of law enforcement officers."

The "Applications" of speaker classification are addressed in the following part: Felix Burckhardt et al. outline scenarios from the area of telephone-based dialog systems. Michael Jessen provides an overview of practical tasks of speaker classification in forensic phonetics and acoustics covering dialect, foreign accent, sociolect, age, gender, and medical conditions. Joaquin Gonzalez-Rodriguez and Daniel Ramos point out an upcoming paradigm shift in the forensic field where the need for objective and standardized procedures is pushing forward the use of automatic speaker recognition methods. Finally, Judith Markowitz sheds some light on the role of speaker classification in the context of the deeper explored sub-fields of speaker recognition and speaker verification.

The next part is concerned with "Methods and Features" for speaker classification beginning with an introduction of the use of frame-based features by Stefan Schacht et al. Higher-level features, i.e., features that rely on either linguistic or long-range prosodic information for characterizing individual speakers are subsequently addressed by Liz Shriberg. Jacques Koreman and his co-authors introduce an approach for enhancing the between-speaker differences at the feature level by projecting the original frame-based feature space into a new feature space using multilayer perceptron networks. An overview of "the features, models, and classifiers derived from [...] the areas of speech science for speaker characterization, pattern recognition and engineering" is provided by Douglas Sturim et al., focusing on the example of modern automatic speaker recognition systems. Izhak Shafran addresses the problem of fusing multiple sources of information, examining in particular how acoustic and lexical information can be combined for affect recognition.

The final part of this volume covers contributions on the "Evaluation" of speaker classification systems. Alvin Martin reports on the last 10 years of speaker recognition evaluations organized by the National Institute for Standards and Technology (nist), discussing how this internationally recognized series of performance evaluations has developed over time as the technology itself has been improved, thereby pointing out the "key factors that have been studied for their effect on performance, including training and test durations, channel variability, and speaker variability." Finally, an evaluation measure which averages the detection performance over various application types is introduced by David van Leeuwen and Niko Brümmer, focusing on its practical applications.

Volume II compiles a number of selected self-contained papers on research projects in the field of speaker classification. The highlights include: Nobuaki Minematsu and Kyoko Sakuraba's report on applying a gender recognition system to estimate the "feminity" of a client's voice in the context of a voice therapy of a “gender identity disorder”; a paper about the effort of studying emotion recognition on the basis of a “real-life” corpus from medical emergency call centers by Laurence Devillers and Laurence Vidrascu; Charl van Heerden and Etienne Barnard’s presentation of a text-dependent speaker verification using features based on the temporal duration of context-dependent phonemes; Jerome Bellegarda’s description of his approach on speaker classification which leverages the analysis of both speaker and verbal content information – as well as studies on accent identification by Emmanuel Ferragne and François Pellegrino, by Mark Huckvale and others.

February 2007 Christian Müller

Table of Contents

I Fundamentals

II Characteristics

III Applications

IV Methods and Features

V Evaluation